

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods

GenRev: Exploring functional relevance of genes in molecular networks

Siyuan Zheng^a, Zhongming Zhao^{a,b,c,*}^a Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA^b Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA^c Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

ARTICLE INFO

Article history:

Received 8 June 2011

Accepted 20 December 2011

Available online 29 December 2011

Keywords:

Gene ranking

Network

Subnetwork

Klein–Ravi algorithm

Limited kWalks algorithm

Disease genes

ABSTRACT

We introduce GenRev, a network-based software package developed to explore the functional relevance of genes generated as an intermediate result from numerous high-throughput technologies. GenRev searches for optimal intermediate nodes (genes) for the connection of input nodes via several algorithms, including the Klein–Ravi algorithm, the limited kWalks algorithm and a heuristic local search algorithm. Gene ranking and graph clustering analyses are integrated into the package. GenRev has the following features. (1) It provides users with great flexibility to define their own networks. (2) Users are allowed to define each gene's importance in a subnetwork search by setting its score. (3) It is standalone and platform independent. (4) It provides an optimization in subnetwork search, which dramatically reduces the running time. GenRev is particularly designed for general use so that users have the flexibility to choose a reference network and define the score of genes. GenRev is freely available at <http://bioinfo.mc.vanderbilt.edu/GenRev.html>.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

High-throughput technologies have enabled researchers to explore a large variety of biological and biomedical problems at the genome-wide scale and have generated huge amounts of biological data [1–4]. These technologies include microarrays (e.g., gene expression, copy number variation, genome-wide association studies, microRNA, and methylation), next generation sequencing (e.g., RNA-Seq, whole exome sequencing, and whole genome sequencing), ChIP-on-CHIP and ChIP-Seq, and proteomics-based platforms. Analyses of the data generated from these technologies often result in a list of noteworthy genes that are useful for biological interpretation and follow up validation [5–9]. Thus, interpretation of gene list has become an important downstream analysis task. To meet this rapidly growing demand, appropriate bioinformatics tools must be developed [10].

Recently, network approaches have been utilized to interpret data gleaned from genomic experiments [11–13]. In particular, it is highly desirable to extract meaningful small subnetworks from the entirety of the reference network. Such subnetworks may disclose gene relationships in the whole network as well as the cooperative signals present in the genomic experiment data. So far, many approaches have been proposed to find these subnetworks. These approaches can generally be separated into two categories: responsive

subnetwork identification and subnetwork extraction initiated by seed genes (or nodes). For the first category, several algorithms and tools are developed by integrating genome-wide measurements of signals (e.g., gene expression from microarrays) with pre-defined networks. Examples include the seminal work by Ideker et al. [12], COSINE by Ma et al. [14], GXNA by Nacu et al. [15], heinz (heaviest induced subgraph) by Dittrich et al. [16], and BioNet by Beisser et al. [17]. These methods typically have a score function and a search strategy and aim to identify the subnetworks with maximal scores in the reference network. In the second category, algorithms typically start with a set of genes as seeds to expand and extract a subnetwork from the reference network. The resultant subnetworks, which reflect the paths in which the seeds are involved [18,19], suggest the functional relationships of the seed genes and further predict additional genes that may play important roles in functional cooperation [20,21].

For subnetwork extraction initiated by seed genes (e.g., gene list of interest), there have been several web-based tools that implement various algorithms. For example, NeAT integrates five algorithms to predict metabolic pathways in its toolkit [18,22]. Genes2Networks uses a neighborhood based approach to connect seed genes in protein interaction networks [23]. While these tools are useful for investigators with specific purposes, each of them has some limitations. (1) Web services essentially rely on internet connections. For a large list of genes, these tools may have a long responding time or their response time may not even be feasible. (2) Genes2Networks uses preloaded networks for the query; thus, it limits the flexibility for those users who need to search subnetworks within their in-house network data. (3) NeAT is designed for metabolic networks,

* Corresponding author at: Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 West End Avenue, Suite 600, Nashville, TN 37203, USA. Fax: +1 615 936 8545.

E-mail address: zhongming.zhao@vanderbilt.edu (Z. Zhao).

so it is not well fitted to other types of network analysis. So far, there is still a lack of a user-friendly tool for a general purpose — extracting subnetworks from the whole network for any list of genes based on user's interest.

Here, we present GenRev, a standalone application to explore the relevance of genes in the context of their networks. Given a reference network and a set of input seed genes, GenRev maps the genes to the network and extracts subnetworks that connect these genes through the use of specific algorithms. The reference network serves as a resource of gene relevance in which nodes typically represent genes and edges represent interactions (or associations). The resulting subnetworks highlight the participant paths and additional nodes connecting to the input genes. Further topological analysis of these subnetworks may help users to identify interesting genes for network structure maintenance. GenRev has the following important features. (1) It provides users with great flexibility to define their networks. (2) Users are allowed to define each gene's importance in a subnetwork search by setting its score. (3) The results are formatted for visualization in the popular network analysis platform Cytoscape. (4) It is standalone and platform independent. (5) It provides an optimization in a subnetwork search, which dramatically reduces the running time and makes the randomization process practical. These features enable GenRev to be suitable for various high-throughput technologies that produce quantitative measures of genes, including genes from next generation sequencing projects. GenRev is freely available at <http://bioinfo.mc.vanderbilt.edu/GenRev.html>.

2. Algorithms and test data

Three algorithms are implemented in GenRev: the Klein–Ravi algorithm [24], the limited kWalks algorithm [25], and a heuristic local search algorithm [11]. Here, we briefly introduce these algorithms. More details of each can be found in the original work (see below) and in the GenRev document (<http://bioinfo.mc.vanderbilt.edu/GenRev/document.pdf>).

2.1. Klein–Ravi algorithm

The algorithm by Klein and Ravi [24] was proposed to solve the node-weighted Steiner tree problem. The goal in the node-weighted Steiner tree problem is to find a subnetwork with a minimum score that could connect all the seeds. In this problem, “seed” is also called “terminal.” The score of a subnetwork is calculated by the sum of the scores of its nodes. The Klein–Ravi algorithm uses a greedy search strategy to connect the seeds. Initially, each seed is set to be a tree by itself. Then, the algorithm iteratively merges the trees to create integrative, larger trees until there is only one tree connecting all seeds. In GenRev, aside from the original greedy strategy in the Klein–Ravi algorithm, we also slightly modified the initialization process of the algorithm. Instead of setting each terminal as an independent tree, we first map terminals to the reference network to check if they have any direct interactions. If some terminals can form a connected graph, then the graph will be used as an initial tree.

In each iteration, the algorithm selects a non-tree node and a subset of at least two trees to minimize the following ratio

$$\frac{\text{node cost} + \text{sum of distance to trees}}{\text{number of trees}}.$$

In most biological and biomedical studies, genes are often scored in proportion to their properties of interest; if such an approach is used, this scoring theme does not immediately fit the Klein–Ravi algorithm. For example, a larger fold change of gene expression indicates a stronger probability of real functional relevance to the phenotypic differences. For this reason, a transformation is needed before we can apply any algorithm. To this end, GenRev internally

transforms the gene score from the user's input into gene (node) cost using the following formula:

$$\text{gene cost} = 1/\sqrt{\text{gene score}}.$$

This algorithm was first introduced to the computational biology field by Scott et al. [19], who demonstrated the algorithm could identify the known yeast regulatory elements of heat shock response as well as gluconeogenesis, galactose, glycolysis and glucose fermentation pathways.

2.2. Limited kWalks algorithm

The limited kWalks algorithm models random walks in networks using a Markov chain and builds a relevant subnetwork that connects seed nodes [18,25]. The relevance of an edge and a node in relation to the seed genes is evaluated by the expected times random walk passes starting from one seed to any of the others. In the interpretation of a graph as a Markov chain, each node represents a state, and the probability of transition from state i to j is given by

$$P_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$$

where w_{ij} is the weight for the edge $i-j$. In GenRev, the users are allowed to set edge scores (weights). By default, all edges have equal weight. More details of the mathematics are available in the GenRev document and the original work by Dupont et al. [25].

2.3. Heuristic local search algorithm

This algorithm is based on a modification of a previously published method [11]. It uses a local expansion approach to find the highest scored neighborhood genes connected with the seeds. In GenRev, seed genes are first mapped to the reference network, and then, each disjoint component of the node-induced network is later used as a seed graph for expansions. A subnetwork score is defined as the sum of its node scores. A search distance d and a minimum score increment rate r is defined to constrain the expansions. Specifically, parameter d defines the maximally distant neighbor nodes being considered in seed graph expansion (e.g., $d=2$ means only nodes with distance 1 or 2 would be considered in expansion), and parameter r defines the threshold value of the score increment after adding each new node. That is, adding a new node to the subnetwork must increase the score by a rate more than r . In GenRev, r is calculated by

$$r = \frac{v_i}{v_i + S_{net}}$$

where v_i is the node score and S_{net} is the seed graph score. By its definition, the range of r is (0, 1).

The expansion is an iteration process. For a seed graph with a score S_{net} , GenRev searches for the maximally scored neighbor within the shortest path d . If the node score is larger than $S_{net} \times (\frac{1}{1-r})$, the addition is valid. Seed graph is then expanded by connecting this node through the shortest path. The iteration terminates when no node satisfies the predefined constraint parameter r .

2.4. Test data

To test GenRev, we used a microarray gene expression data set [26] and the whole human protein–protein interaction (PPI) network from the Protein Interaction Network Analysis (PINA) platform [27]. This microarray dataset collected 75 hepatitis C virus (HCV) infected hepatocellular carcinoma (HCC) samples, cataloguing 8 disease stages. In our testing, we broadly divided these samples into a

cancerous group and a precancerous group, resulting in 35 and 37 samples respectively. For data integrity, we excluded 3 samples from cirrhotic liver tissue from patients without HCC. The microarrays were normalized by the MAS algorithm as done in the previous work [26]. Signals at the probeset level were collapsed to the gene level by using the strongest value in each sample. Genes were scored and ranked by their fold changes (logarithm 2 scale, absolute value) between the two groups. The entity IDs of PPI data was converted to gene symbols using the *org.Hs.eg.db* package in Bioconductor (version 2.5). All test data, as well as reference networks (human protein–protein interaction network prepared from the PINA database [27], yeast functional network [28], networks available at VisANT [29]), can be downloaded at the GenRev website (<http://bioinfo.mc.vanderbilt.edu/GenRev.html>).

3. Results

3.1. Implementation of algorithms

GenRev is implemented in the Python programming language with additional packages NetworkX [30] and NumPy [21]. It is a command-line standalone application. Each algorithm is implemented in one module. A community structure (also called modules in networks) detection algorithm, the Markov Cluster Algorithm (MCL) [31], is implemented to analyze the resulting subnetworks. An indicative measurement, modularity, is computed for module detection according to the original definition in Newman [32]. Additionally, GenRev provides an option for the users to apply a network pruning strategy to dramatically reduce the computational time (see Section 3.2 below). To assess whether the resultant subnetworks are from the whole network by chance, we provide R scripts to generate random gene sets and run the subnetwork extraction using the same algorithm and parameters. Network randomization analysis typically requires the users to run 1000 random sets to evaluate the significance of the resultant subnetworks. We provided two R scripts, one for node weight randomization [11] and another for edge-based

randomization [33]. These two R scripts are available on the GenRev website. Finally, GenRev does not provide a function for network visualization. Instead, it outputs files in a format compatible with the popular visualization tool Cytoscape [34]. Accordingly, users can export the results to Cytoscape for network visualizations, including generation of figures for presentation and publication purposes. Fig. 1 illustrates the flowchart of GenRev implementation.

3.2. Runtime test

An approach to effectively shorten the runtime of a subnetwork search is to reduce the reference network size while retaining as much information as possible. By default, GenRev applies a simple yet efficient network pruning strategy that can be controlled by the users. Briefly, users can set a path length s , and then, the network induced by the nodes within the length s of the seeds can be used as the pruned reference network. Assuming that most genes in the input list are coherently associated (e.g., co-deregulated in the experiments, all related to a common disease), they likely have a short distance in the network. Thus, this strategy can dramatically increase the efficiency in a subnetwork search while retaining most interaction relationships among the input genes.

Using the test data in Section 2.4, we tested the functions of GenRev utilizing several sets of seeds. The testing server runs on a Linux system, with an Intel core 2, 3.0 GHz CPU and 16 GB memory. As shown in Table 1, GenRev was able to run very efficiently. For example, it only took 68 s to rank the top 100 genes using the network pruning strategy of the Klein–Ravi algorithm implemented in GenRev. However, if we use the same algorithm without pruning, it would require ~2.5 h. Additionally, we assessed the computational time in the randomization process. It took 30 s to analyze the top 200 genes from the test data using the heuristic local search algorithm implemented in GenRev. We used both R scripts to perform 1000 randomization analyses for the same set of genes, reference network, algorithm, and parameters. We ran each script through five threads in parallel. For the edge-based randomization, it took

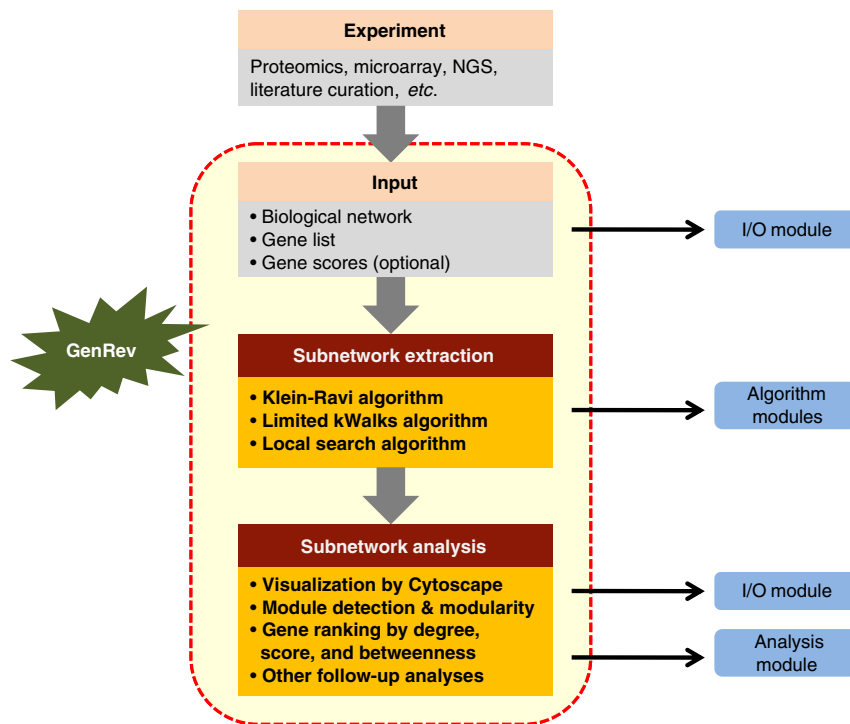


Fig. 1. Workflow of GenRev. GenRev can be broadly divided into three components: I/O module, algorithm modules, and analysis module. The I/O module employs functions to read input data and write results to the output directory. Algorithm modules implement the subnetwork search algorithms. The analysis module is utilized for subnetwork analyses such as community structure detection and gene ranking.

Table 1

GenRev runtime with different sets of seed gene lists.

| Algorithm | Top 50 genes | Top 100 genes | Top 200 genes |
|------------------------------|--------------|---------------|---------------|
| Klein–Ravi – pruning | 7 s | 68 s | 8 min |
| Klein–Ravi – without pruning | 4 h 24 min | 2 h 26 min | 1 h 41 min |
| Limited kWalks | 25 s | 8 min | 42 min |
| Heuristic local search | 7 s | 13 s | 30s |

S: seconds; min: minute; h: hour.

18,015, 19,478, 17,802, 18,171, and 19,042 s for each of these threads and approximately 5.4 h total to complete the whole randomization analysis. Similarly, for the node-based randomization, it took 14,824, 15,245, 14,745, 14,702, and 14,230 s for each thread and approximately 4.2 h total to complete the whole randomization analysis.

3.3. Previous applications

Algorithms implemented in GenRev have been used in several studies. For example, Scott et al. [19] used the Klein–Ravi algorithm applied to a yeast regulatory network. Querying a list of differentially

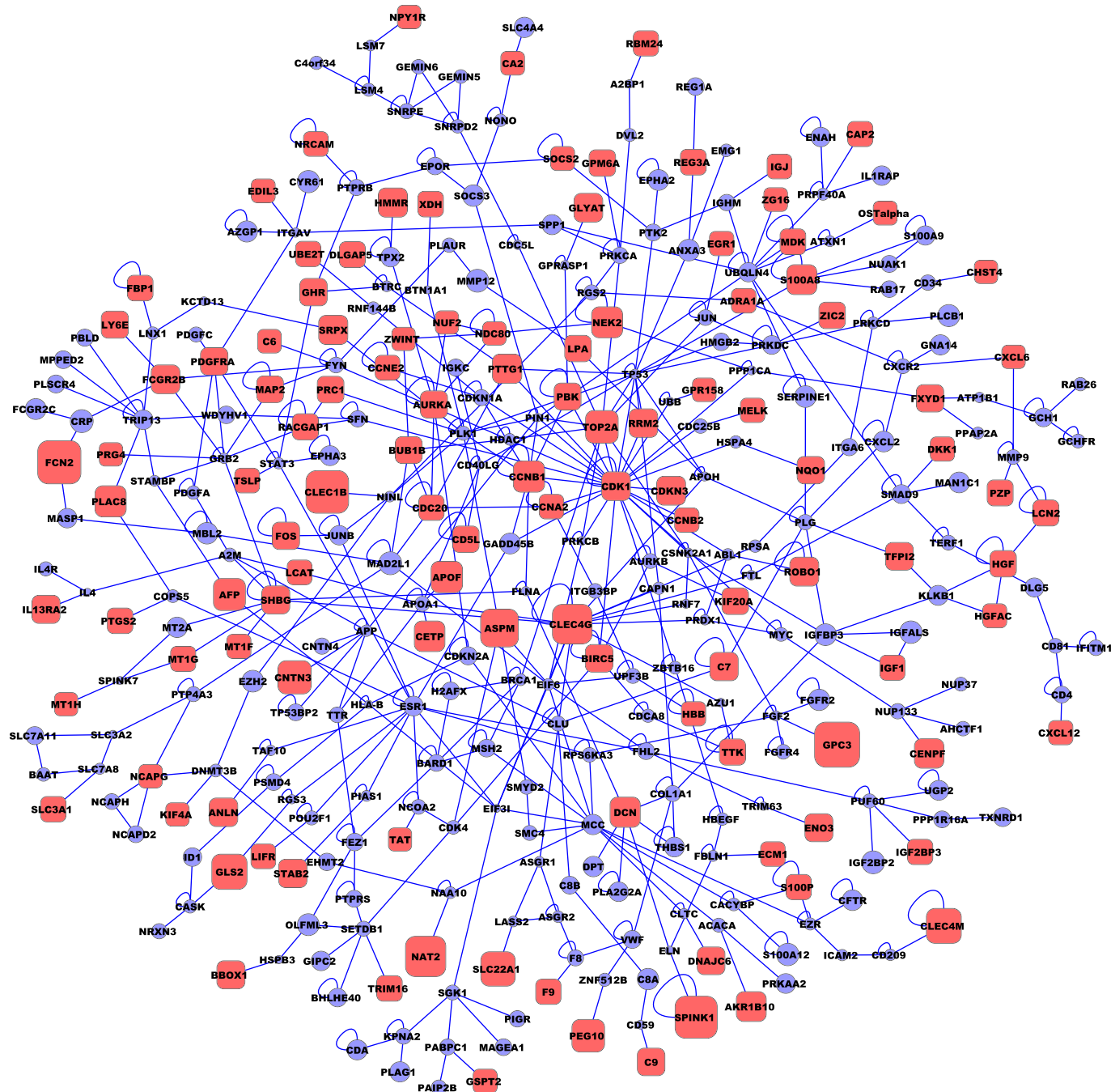


Fig. 2. The subnetwork identified from the protein–protein interaction network using the heuristic local search algorithm in GenRev. In this test data, we used 200 differentially expressed genes from a microarray expression experiment on hepatitis C virus (HCV) infected hepatocellular carcinoma (HCC) samples. Each node represents a protein, and each edge represents the physical interactions between the two end nodes. Nodes in red denote input seeds, and nodes in blue denote recruited nodes connecting the seeds by the algorithm. The size of a node is proportional to its score.

expressed genes and GAL80, they found a subnetwork connecting GAL80 and diauxic shift. They also found general transcription factors RAP1 and HSF1 regulated transcriptional programs in yeast. A heuristic local search algorithm, which is similar to that implemented in GenRev, has been applied to search subnetworks with enriched signals from genome-wide association studies (GWAS) [13]. In that study, each gene from the GWAS dataset was considered a seed, and its network module was expanded based on local association signals. A randomization process (100,000 times) was performed to evaluate whether the module was generated by chance.

We applied an early version of GenRev to several psychiatric genetic studies. Applying the Klein–Ravi algorithm, we extracted a schizophrenia specific network and a cancer specific network using disease related gene lists [20]. Moreover, we explored the network properties and compared them [20]. We further validated novel candidate genes from the schizophrenia subnetwork [35]. In another example, we used 373 epilepsy genes from the copy number variation (CNV) regions and 165 epilepsy genes from the HuGE Navigator (<http://hugenavigator.net/HuGENavigator/>) as seeds and extracted two epilepsy specific subnetworks. Subsequently, 20 genes were prioritized as epilepsy candidate genes. Two of them, *CHRNA7* and *GABRA1*, were further evaluated as differentially expressed genes in an epilepsy expression data set [21].

3.4. New case study

To further illustrate the capability of GenRev, we applied the heuristic local search algorithm to the test data (HCC data in Section 2.4) with default parameters. We use the top 200 differentially expressed genes as seeds. The results are shown in Fig. 2 (the figure is available in high resolution at <http://bioinfo.mc.vanderbilt.edu/GenRev/hcc.pdf>). Other sets of seed genes were also tested, and all results are available on the GenRev website. Interestingly, a cell cycle regulatory module was identified through the use of all settings, highlighting the dysregulation of cell cycle homeostasis in HCV induced HCC. Genes in this module include *CDK1*, *BIRC5*, *GADD45B*, etc. *CDK1* is the hub of the cell cycle module in the subnetworks, suggesting that functional blocking of its activity may trigger a systematic repression of tumor progression, as validated in some other tumors too [36]. The well known tumor suppressor gene *TP53* was also highlighted as one whose encoded protein is a hub in the subnetwork. Another interesting gene revealed by GenRev ranking is *CLEC4G*, which was prioritized due to high degree and betweenness. Recent studies reported that the protein encoded by this gene could interact with two HCV receptors [37], and, moreover, it regulates hepatic T-cell immune responses [38]. Our results suggest the proteins encoded by those genes have strong functional relationship in the local context, which might facilitate the generation of new hypotheses.

In another study, we applied the Klein–Ravi algorithm in GenRev to our recently collected genes for Alzheimer's disease (AD) and Parkinson's disease (PD). The comparison of the extracted AD- and PD-specific subnetworks from the whole human interactome revealed substantial overlap, suggesting a strong connection between the two diseases at the molecular network level [39].

4. Conclusion

Genes do not function in isolation; rather, they interact extensively with other molecules within their molecular networks [40,41]. With the increased availability of genome-wide experiment technologies, more and more disease- or phenotype-associated genes will be identified, thereby supplying a wealth of information in order to understand the biological groundwork of observed consequences. Meanwhile, deciphering this massive amount of information is challenging. In this work, we present an open source network tool, GenRev, to interpret the genes that result from genomic experiments

in the context of molecular networks. Compared to other network tools such as NeAT and Genes2Networks, GenRev is uniquely designed for a general purpose approach and can be used for any list of genes. Users can choose any form of reference network, and the node score can be generated from any genomic experiment. These features position GenRev as a useful downstream analysis tool when facing a deluge of data from microarrays (gene expression, CNVs, microRNA, methylation, etc.), next generation sequencing, proteomics, literature mining, and multiple dimensional data integration, among other high-throughput data sets. Future development of GenRev will focus on the assessment of the significance of the resultant subnetworks, their biological interpretation, and integrating additional biology knowledge in gene prioritization.

Acknowledgments

This project was partially supported by the NIH grants (R03DE022093) and 2009 NARSAD Maltz Investigator Award (to Z.Z.). We thank the anonymous reviewers for their valuable comments. We thank Dr. Peilin Jia for her valuable contribution in the revision of the manuscript, Dr. Pierre Dupont for his kind help in the implementation of the limited k-walk algorithm, and Ms. Rebecca H. Posey for critically reading and improving the manuscript. We also thank the members of the Bioinformatics and Systems Medicine Laboratory in Vanderbilt University for their useful discussion and suggestions.

References

- [1] T. Pastinen, Genome-wide allele-specific analysis: insights into regulatory variation, *Nat. Rev. Genet.* 11 (2010) 533–538.
- [2] C.G. Liu, R. Spizzo, G.A. Calin, C.M. Croce, Expression profiling of microRNA using oligo DNA arrays, *Methods* 44 (2008) 22–30.
- [3] T. Ideker, V. Thorsson, J.A. Raniish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, L. Hood, Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science* 292 (2001) 929–934.
- [4] X. Zhang, J. Yazaki, A. Sundaresan, S. Cokus, S.W. Chan, H. Chen, I.R. Henderson, P. Shinn, M. Pellegrini, S.E. Jacobsen, et al., Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis, *Cell* 126 (2006) 1189–1201.
- [5] P. Hao, S. Zheng, J. Ping, K. Tu, C. Gieger, R. Wang-Sattler, Y. Zhong, Y. Li, Human gene expression sensitivity according to large scale meta-analysis, *BMC Bioinforma.* 10 (Suppl 1) (2009) S56.
- [6] S. Anders, W. Huber, Differential expression analysis for sequence count data, *Genome Biol.* 11 (2010) R106.
- [7] W.E. Johnson, W. Li, C.A. Meyer, R. Gottardo, J.S. Carroll, M. Brown, X.S. Liu, Model-based analysis of tiling-arrays for ChIP-chip, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 12457–12462.
- [8] S.T. Kim, K.S. Cho, S. Yu, S.G. Kim, J.C. Hong, C.D. Han, D.W. Bae, M.H. Nam, K.Y. Kang, Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells, *Proteomics* 3 (2003) 2368–2378.
- [9] A. Adeyemo, N. Gerry, G. Chen, A. Herbert, A. Doumatey, H. Huang, J. Zhou, K. Lashley, Y. Chen, M. Christman, et al., A genome-wide association study of hypertension and blood pressure in African Americans, *PLoS Genet.* 5 (2009) e1000564.
- [10] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* 4 (2009) 44–57.
- [11] H.Y. Chuang, E. Lee, Y.T. Liu, D. Lee, T. Ideker, Network-based classification of breast cancer metastasis, *Mol. Syst. Biol.* 3 (2007) 140.
- [12] T. Ideker, O. Ozier, B. Schwikowski, A.F. Siegel, Discovering regulatory and signalling circuits in molecular interaction networks, *Bioinformatics* 18 (Suppl 1) (2002) S233–S240.
- [13] P. Jia, S. Zheng, J. Long, W. Zheng, Z. Zhao, dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks, *Bioinformatics* 27 (2011) 95–102.
- [14] H. Ma, E.E. Schadt, L.M. Kaplan, H. Zhao, COSINE: COndition-Specific sub-NetWork identification using a global optimization method, *Bioinformatics* 27 (2011) 1290–1298.
- [15] S. Nacu, R. Critchley-Thorne, P. Lee, S. Holmes, Gene expression network analysis and applications to immunology, *Bioinformatics* 23 (2007) 850–858.
- [16] M.T. Dittrich, G.W. Klau, A. Rosenwald, T. Dandekar, T. Muller, Identifying functional modules in protein–protein interaction networks: an integrated exact approach, *Bioinformatics* 24 (2008) i223–i231.
- [17] D. Beisser, G.W. Klau, T. Dandekar, T. Muller, M.T. Dittrich, BioNet: an R-Package for the functional analysis of biological networks, *Bioinformatics* 26 (2010) 1129–1130.
- [18] K. Faust, P. Dupont, J. Callut, J. van Helden, Pathway discovery in metabolic networks by subgraph extraction, *Bioinformatics* 26 (2010) 1211–1218.

- [19] M.S. Scott, T. Perkins, S. Bunnell, F. Pepin, D.Y. Thomas, M. Hallett, Identifying regulatory subnetworks for a set of genes, *Mol. Cell. Proteomics* 4 (2005) 683–692.
- [20] J. Sun, P. Jia, A.H. Fanous, E. van den Oord, X. Chen, B.P. Riley, R.L. Amdur, K.S. Kendler, Z. Zhao, Schizophrenia gene networks and pathways and their applications for novel candidate gene selection, *PLoS One* 5 (2010) e11351.
- [21] P. Jia, J.M. Ewers, Z. Zhao, Prioritization of epilepsy associated candidate genes by convergent analysis, *PLoS One* 6 (2011) e17162.
- [22] S. Brohee, K. Faust, G. Lima-Mendez, G. Vanderstocken, J. van Helden, Network analysis tools: from biological networks to clusters and pathways, *Nat. Protoc.* 3 (2008) 1616–1629.
- [23] S.I. Berger, J.M. Posner, A. Ma'ayan, Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases, *BMC Bioinforma.* 8 (2007) 372.
- [24] P. Klein, R. Ravi, A nearly best-possible approximation algorithm for node-weighted Steiner trees, *J. Algorithms* 19 (1995) 104–115.
- [25] P. Dupont, J. Callut, G. Dooms, J.N. Monette, Y. Deville, Relevant subgraph extraction from random walks in a graph, Research report UCL/FSA/INGI 2006–07, 2006.
- [26] E. Wurmbach, Y.B. Chen, G. Khitrov, W. Zhang, S. Roayaie, M. Schwartz, I. Fiel, S. Thung, V. Mazzaferro, J. Bruix, et al., Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma, *Hepatology* 45 (2007) 938–947.
- [27] J. Wu, T. Vallenius, K. Ovaska, J. Westermarck, T.P. Makela, S. Hautaniemi, Integrated network analysis platform for protein–protein interactions, *Nat. Methods* 6 (2009) 75–77.
- [28] I. Lee, S.V. Date, A.T. Adai, E.M. Marcotte, A probabilistic functional network of yeast genes, *Science* 306 (2004) 1555–1558.
- [29] Z. Hu, J.H. Hung, Y. Wang, Y.C. Chang, C.L. Huang, M. Huyck, C. DeLisi, VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology, *Nucleic Acids Res.* 37 (2009) W115–W121.
- [30] A.A. Hagberg, D.A. Schult, P.J. Sieter, Exploring network structure, dynamics, and function using NetworkX, Proceedings of the 7th Python in Science Conference (SciPy2008), 2008, pp. 11–15.
- [31] S.v. Dongen, Graph clustering by flow simulation, PhD thesis, University of Utrecht (2000).
- [32] M.E. Newman, Fast algorithm for detecting community structure in networks, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69 (2004) 066133.
- [33] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, *Science* 298 (2002) 824–827.
- [34] M.S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, et al., Integration of biological networks and gene expression data using Cytoscape, *Nat. Protoc.* 2 (2007) 2366–2382.
- [35] J. Sun, C. Wan, P. Jia, A.H. Fanous, K.S. Kendler, B.P. Riley, Z. Zhao, Application of systems biology approach identifies and validates GRB2 as a risk gene for schizophrenia in the Irish Case Control Study of Schizophrenia (ICSS) sample, *Schizophr. Res.* 125 (2011) 201–208.
- [36] A. Goga, D. Yang, A.D. Tward, D.O. Morgan, J.M. Bishop, Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC, *Nat. Med.* 13 (2007) 820–827.
- [37] Y. Li, B. Hao, X. Kuai, G. Xing, J. Yang, J. Chen, L. Tang, L. Zhang, F. He, C-type lectin LSECtin interacts with DC-SIGNR and is involved in hepatitis C virus binding, *Mol. Cell. Biochem.* 327 (2009) 183–190.
- [38] L. Tang, J. Yang, W. Liu, X. Tang, J. Chen, D. Zhao, M. Wang, F. Xu, Y. Lu, B. Liu, et al., Liver sinusoidal endothelial cell lectin, LSECtin, negatively regulates hepatic T-cell immune response, *Gastroenterology* 137 (2009) 1498–1508 e1491–1495.
- [39] Y. Liu, M. Koyuturk, S. Maxwell, Z. Zhao, M.R. Chance, Integrative analysis of common neurodegenerative diseases using gene association, interaction network and mRNA expression data, American Medical Informatics Association Summit on Translational Bioinformatics Proceedings, 2012, March 19–21, 2012, San Francisco, CA, USA.
- [40] K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.L. Barabasi, The human disease network, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 8685–8690.
- [41] T. Ideker, R. Sharan, Protein networks in disease, *Genome Res.* 18 (2008) 644–652.